

Aqui está uma descrição detalhada para o projeto de Análise de Imagens de Satélite com PCA, formatada como um guia para os alunos.

Roteiro do Projeto: Análise de Imagens de Satélite com PCA

Título: Análise de Componentes Principais (PCA) para Realce de Feições em Imagens de Satélite Multiespectrais 🗺️

Resumo do Projeto

Imagens de satélite modernas capturam a superfície da Terra em múltiplas bandas do espectro eletromagnético (além do visível), que são frequentemente muito correlacionadas entre si. Este projeto consiste em implementar a Análise de Componentes Principais (PCA) do zero para transformar essas bandas correlacionadas em um novo conjunto de bandas não correlacionadas (os componentes principais). O objetivo é demonstrar como essa técnica de Álgebra Linear Numérica pode reduzir a redundância dos dados e realçar feições geográficas, como corpos d'água, tipos de vegetação e áreas urbanas, que não são facilmente distinguíveis nas bandas originais.

Objetivos de Aprendizagem

- **Compreender** a aplicação prática da SVD ou da decomposição em autovalores como uma ferramenta de redução de dimensionalidade e extração de características.
- **Implementar** a pipeline completa da PCA: cálculo da matriz de covariância, resolução do problema de autovalor e projeção dos dados.
- **Aprender** a manipular dados geoespaciais (imagens de satélite em formato raster).
- **Analisar e interpretar** os resultados visuais da transformação, conectando os componentes principais a padrões físicos no terreno.

Conceitos Fundamentais do Curso

- **Decomposição em Autovalores e SVD:** O coração do método.
- **Matriz de Covariância:** Como medir a relação entre múltiplas variáveis.
- **Projeção de Vetores e Mudança de Base:** A essência da transformação dos dados.
- **Operações com Matrizes e Vetores:** A base para toda a implementação.

Roteiro de Implementação e Análise

Fase 1: Obtenção e Preparação dos Dados

1. Download da Imagem:

- Acesse um portal de dados de satélite públicos, como o **USGS EarthExplorer** ou o **Copernicus Open Access Hub**.
- Procure e baixe uma cena do satélite **Landsat 8/9** ou **Sentinel-2** para uma área de interesse que contenha feições diversas (cidade, floresta, água, agricultura). Escolha uma cena com poucas nuvens.
- Você receberá várias imagens em tons de cinza, uma para cada banda espectral

(ex: B2_blue.tif, B3_green.tif, B4_red.tif, B5_nir.tif, etc.).

2. Leitura dos Dados:

- O aluno deve implementar um script em Python para ler as bandas espectrais relevantes (por exemplo, as bandas visíveis e do infravermelho próximo).
- **Ferramentas Sugeridas:** Utilize bibliotecas como rasterio ou GDAL para ler os arquivos GeoTIFF e extrair as matrizes numéricas (arrays NumPy).
- **Empilhamento:** Empilhe as matrizes de cada banda para formar um tensor 3D (largura x altura x número de bandas).

Fase 2: Implementação do Algoritmo PCA

O aluno deve implementar as funções para realizar a PCA, não apenas usar uma biblioteca pronta.

1. Reformatação dos Dados:

- Transforme o tensor 3D em uma matriz 2D. A nova matriz X terá (largura * altura) linhas e (número de bandas) colunas. Cada linha é um pixel e cada coluna é uma banda.

2. Cálculo da Matriz de Covariância:

- Primeiro, normalize os dados subtraindo a média de cada coluna (banda).
- Em seguida, implemente o cálculo da matriz de covariância S para a matriz de dados normalizada. A fórmula é $S = \frac{1}{N-1} X^T X$, onde N é o número de pixels.

3. Resolução do Problema de Autovalor:

- Implemente um algoritmo para encontrar os **autovalores** (λ_i) e **autovetores** (v_i) da matriz de covariância S . Os autovetores v_i são os **Componentes Principais**.
- Ordene os autovalores em ordem decrescente e reordene os autovetores correspondentes.

4. Projeção dos Dados:

- Crie a matriz de transformação P , cujas colunas são os autovetores ordenados.
- Projete os dados originais (normalizados) na nova base dos componentes principais: $Y = XP$. A matriz Y contém os valores dos pixels em cada componente principal.

Fase 3: Análise e Visualização dos Resultados

Esta é a fase de interpretação geoespacial.

1. Reformatar para Imagem:

Transforme a matriz de dados projetada Y de volta para o formato de imagem (uma imagem para cada componente principal).

2. Visualização:

- Plote cada componente principal como uma imagem em tons de cinza.
- Crie uma composição colorida (RGB) usando três componentes principais diferentes (ex: PC1 no canal Vermelho, PC2 no Verde e PC3 no Azul) para criar uma imagem de "falsas-cores" que realça diferentes feições.

3. Análise do Relatório:

- Apresente as imagens de cada componente principal e a composição colorida.
- Calcule a porcentagem da variância total explicada por cada componente (usando os autovalores). Tipicamente, os primeiros 2 ou 3 PCs capturam >95% da

variância.

- **Interprete o que cada componente está mostrando.** Por exemplo, o PC1 geralmente representa o brilho médio ("albedo"). O PC2 pode mostrar um forte contraste entre vegetação e água, pois o infravermelho (que domina esta variância) é fortemente refletido por plantas e absorvido por água.
- Compare a imagem PCA com uma composição RGB colorida "verdadeira" das bandas originais e discuta as vantagens da transformação.

Referências Sugeridas

- **Primária (Teoria NLA):** Trefethen & Bau, "Numerical Linear Algebra", Aulas 4, 5 (SVD) e 26-29 (Eigenvalue Problems). Watkins, "Fundamentals of Matrix Computations", para um tratamento mais rigoroso.
- **Contexto GIS/Remoto:** "Principal Component Analysis in Remote Sensing" - existem muitos artigos e notas de aula com este título. O livro "Remote Sensing of the Environment: An Earth Resource Perspective" por Jensen é uma referência clássica.
- **Dados:** Portal **USGS EarthExplorer** (requer cadastro gratuito).
- **Ferramentas Python:** Tutoriais das bibliotecas rasterio (para leitura de dados), NumPy (para as operações matriciais) e Matplotlib (para visualização).